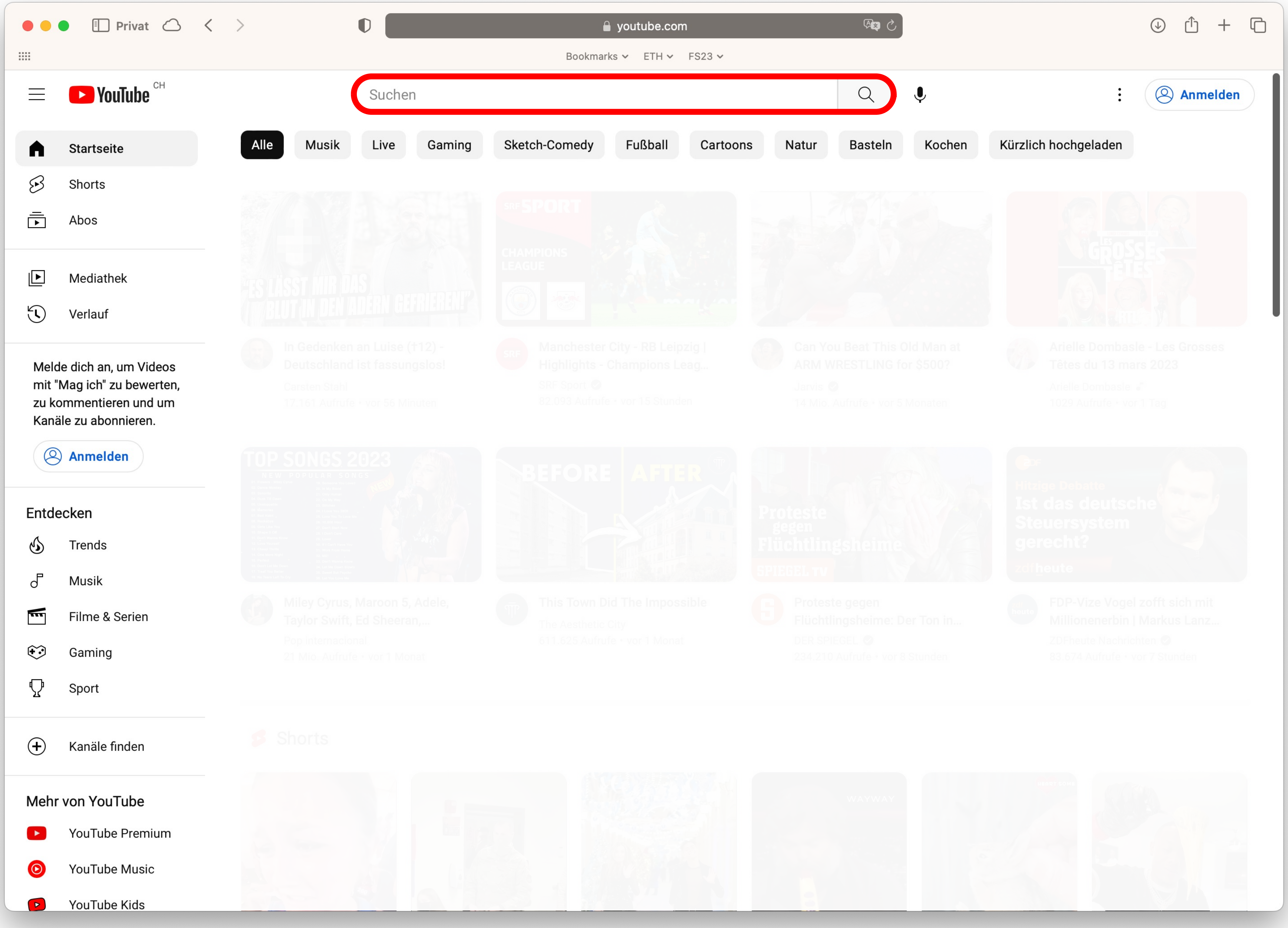
An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes

Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, Maria Bielikova

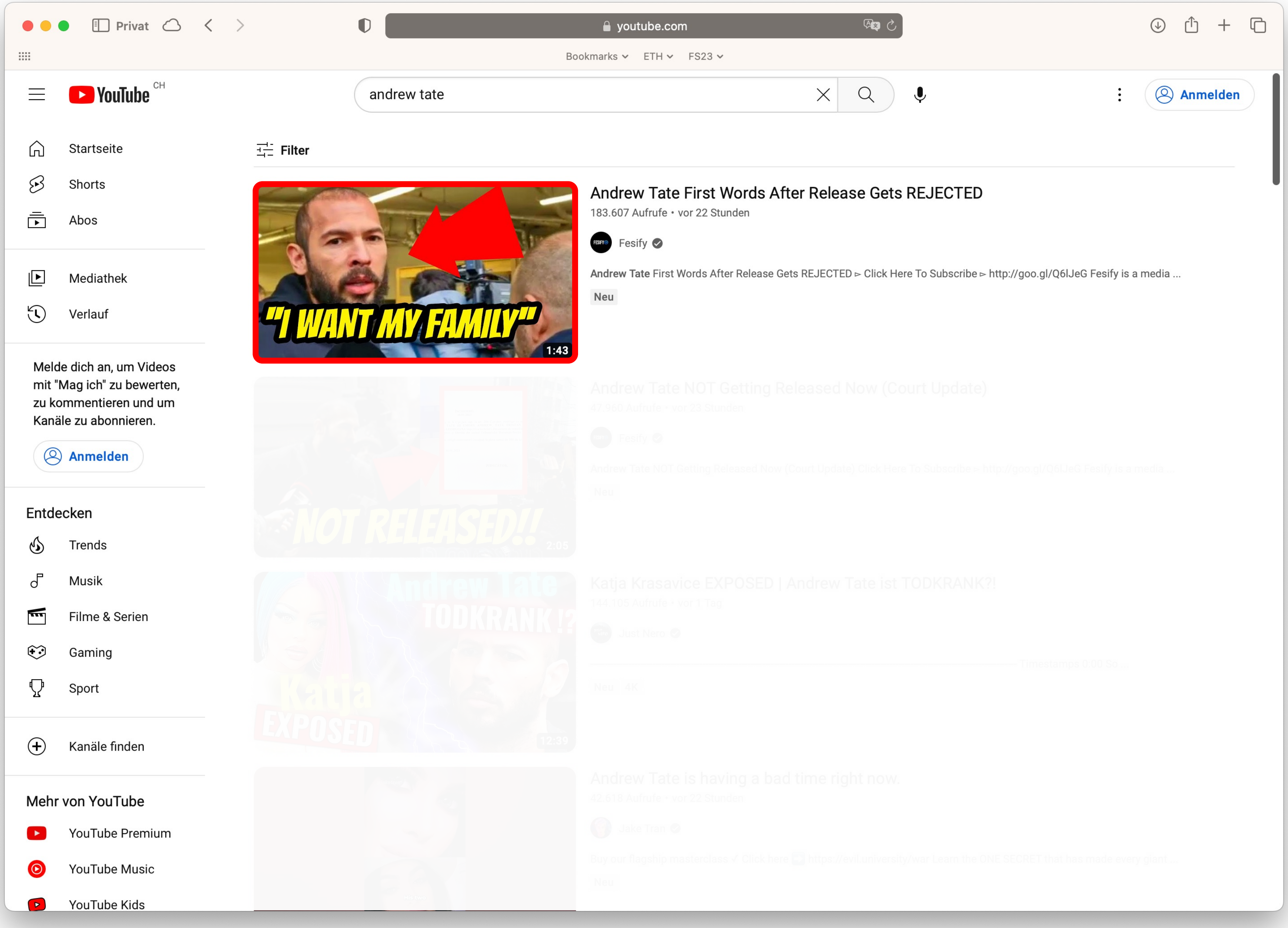
Stuart Heeb

Have you ever noticed…

Let’s take a trip down the YouTube rabbit hole


Privat
youtube.com
Bookmarks
ETH
FS23
YouTube CH
Suchen
Anmelden
Alle
Musik
Live
Gaming
Sketch-Comedy
Fußball
Cartoons
Natur
Basteln
Kochen
Kürzlich hochgeladen
Startseite
Shorts
Abos
Mediathek
Verlauf
Entdecken
Trends
Musik
Filme & Serien
Gaming
Sport
Kanäle finden
Mehr von YouTube
YouTube Premium
YouTube Music
YouTube Kids
In Gedenken an Luise (†12) - Deutschland ist fassungslos!
Manchester City - RB Leipzig | Highlights - Champions Leag...
Can You Beat This Old Man at ARM WRESTLING for $500?
Arielle Dombasle - Les Grosses Têtes du 13 mars 2023
Miley Cyrus, Maroon 5, Adele, Taylor Swift, Ed Sheeran,...
This Town Did The Impossible
Proteste gegen Flüchtlingsheime: Der Ton in...
FDP-Vize Vogel zofft sich mit Millionenerbin | Markus Lanz...
Shorts

Melde dich an, um Videos mit "Mag ich" zu bewerten, zu kommentieren und um Kanäle zu abonnieren.



Andrew Tate First Words After Release Gets REJECTED

183.607 Aufrufe • vor 22 Stunden

Fesify

Andrew Tate First Words After Release Gets REJECTED ▻ Click Here To Subscribe ▻ http://goo.gl/Q6lJeG Fesify is a media ...

Neu

Andrew Tate NOT Getting Released Now (Court Update)

47.960 Aufrufe • vor 23 Stunden

Fesify

Andrew Tate NOT Getting Released Now (Court Update) Click Here To Subscribe ▻ http://goo.gl/Q6lJeG Fesify is a media ...

Neu

Katja Krasavice EXPOSED | Andrew Tate ist TODKRANK?!

144.105 Aufrufe • vor 1 Tag

Just Nero

———————————————————— Timestamps 0:00 So ...

Neu 4K

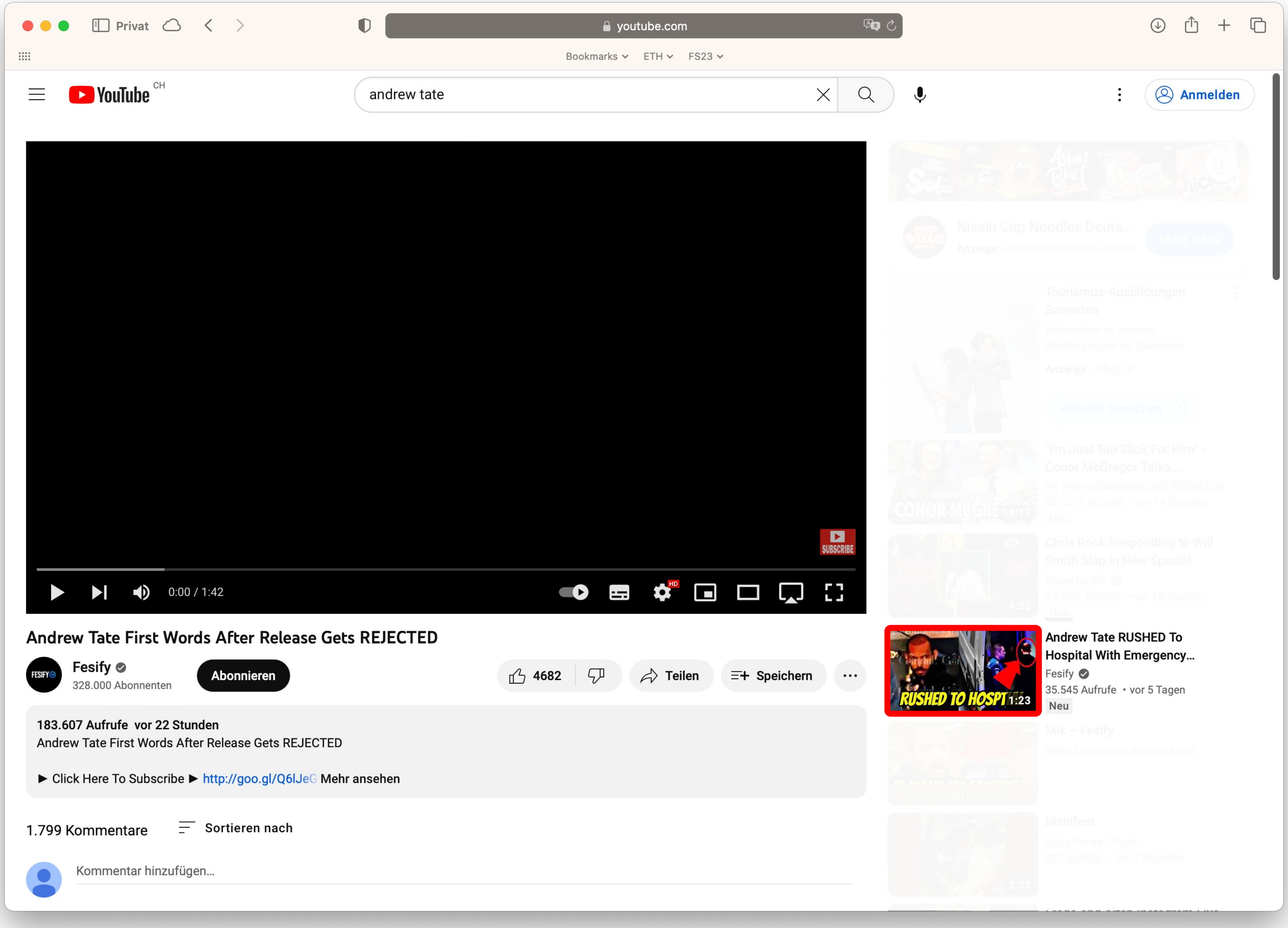
Andrew Tate is having a bad time right now.

42.618 Aufrufe • vor 22 Stunden

Jake Tran

Buy our flagship masterclass ✓ Click here ➡ https://evil.university/war Learn the ONE SECRET that has made every giant ...

Neu



andrew tate

Andrew Tate First Words After Release Gets REJECTED

Fesify
328.000 Abonnenten

4682

183.607 Aufrufe vor 22 Stunden
Andrew Tate First Words After Release Gets REJECTED

▶ Click Here To Subscribe ▶ http://goo.gl/Q6lJeC Mehr ansehen

1.799 Kommentare

Kommentar hinzufügen...

Nissin Cup Noodles Deuts...
Anzeige · de.nissin-foods.eu/cupnoc

Tourismus-Ausbildungen Samedan
Information zu unseren Studiengängen im Tourismus
Anzeige · hftgr.ch

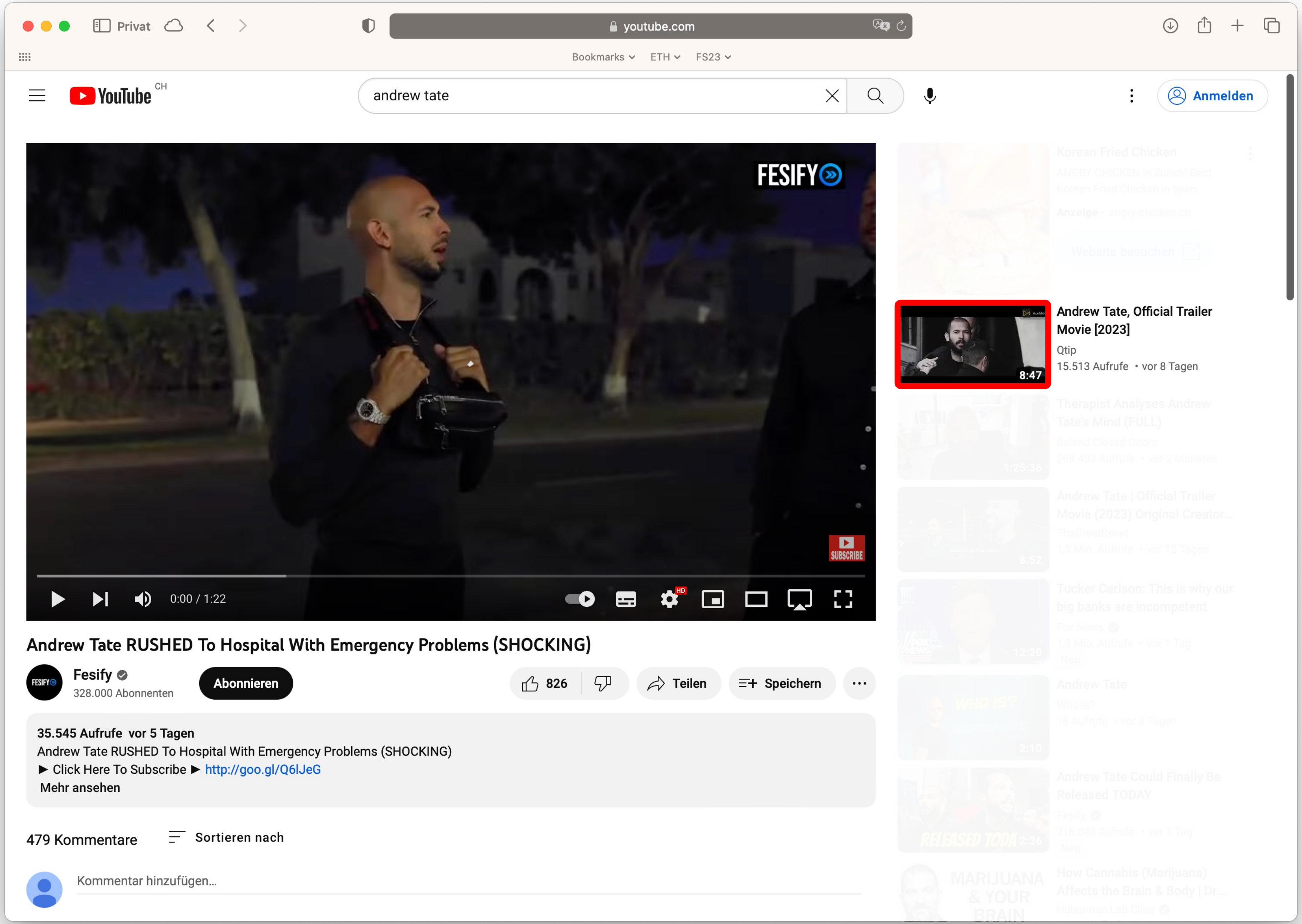
"I'm Just Too Slick For Him" - Conor McGregor Talks...
My Mom's Basement with Robbie Fox
287.214 Aufrufe • vor 14 Stunden
Neu

Chris Rock Responding to Will Smith Slap in New Special
PowerfulJRE
2,2 Mio. Aufrufe • vor 19 Stunden
Neu

Andrew Tate RUSHED To Hospital With Emergency...
Fesify
35.545 Aufrufe • vor 5 Tagen
Neu

Mix – Fesify
Mehr für dich von diesem Kanal

Manifest
Kyyla Renee - Topic
985 Aufrufe • vor 2 Monaten



Andrew Tate RUSHED To Hospital With Emergency Problems (SHOCKING)

Fesify
328.000 Abonnenten

35.545 Aufrufe vor 5 Tagen
Andrew Tate RUSHED To Hospital With Emergency Problems (SHOCKING)
▶ Click Here To Subscribe ▶ http://goo.gl/Q6lJeG
Mehr ansehen

479 Kommentare

Korean Fried Chicken
ANGRY CHICKEN in Zurich! Best Korean Fried Chicken in town.
Anzeige · angry-chicken.ch
Website besuchen

8:47 Andrew Tate, Official Trailer Movie [2023]
Qtip
15.513 Aufrufe • vor 8 Tagen

1:25:36 Therapist Analyses Andrew Tate's Mind (FULL)
Behind Closed Doors
268.493 Aufrufe • vor 2 Monaten

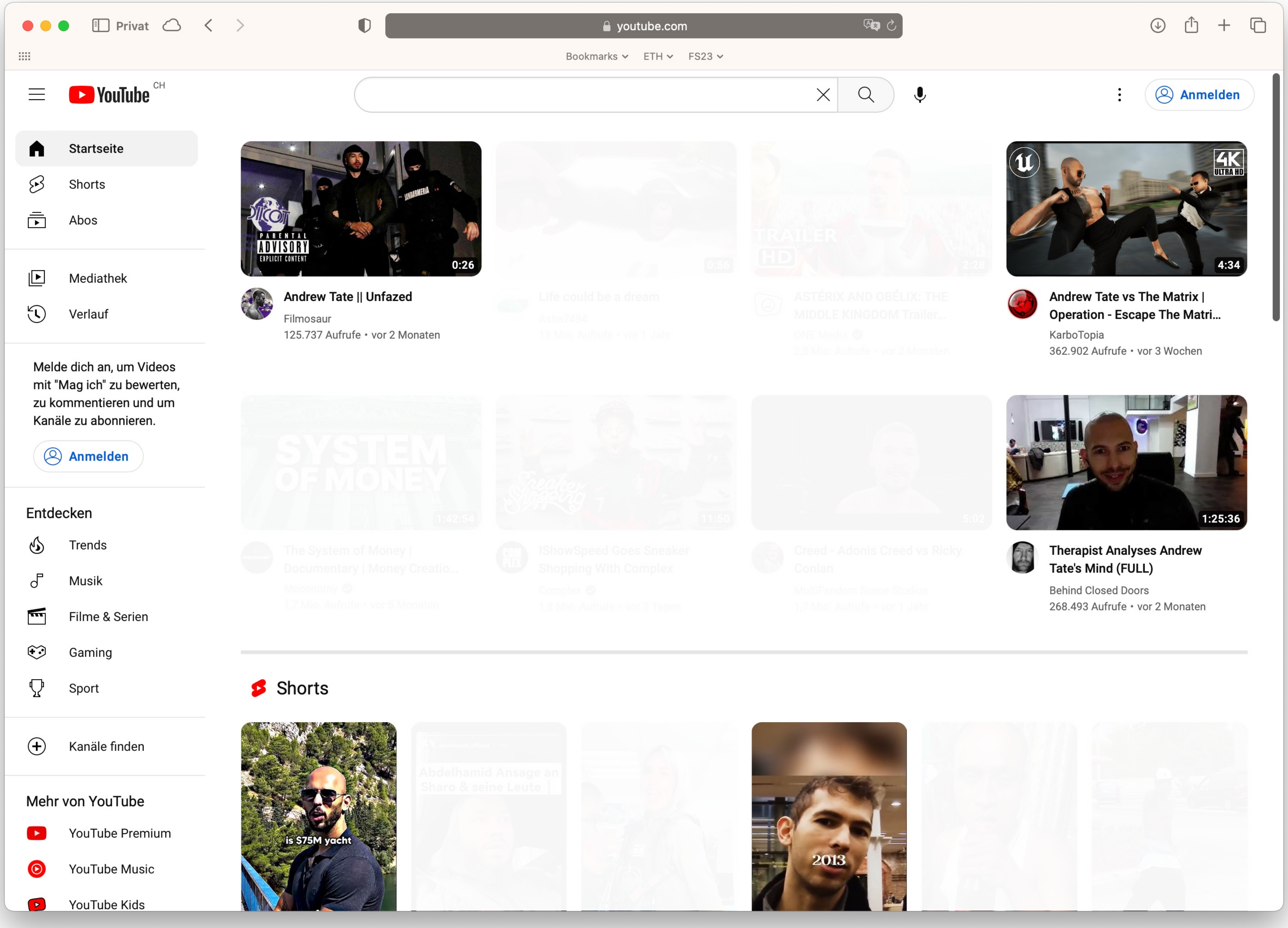
8:52 Andrew Tate | Official Trailer Movie (2023) Original Creator...
ThaGreatReset
1,3 Mio. Aufrufe • vor 12 Tagen

12:20 Tucker Carlson: This is why our big banks are incompetent
Fox News
1,3 Mio. Aufrufe • vor 1 Tag
Neu

2:10 Andrew Tate
Who-Is?
18 Aufrufe • vor 8 Tagen

2:36 Andrew Tate Could Finally Be Released TODAY
Fesify
216.848 Aufrufe • vor 1 Tag
Neu

How Cannabis (Marijuana) Affects the Brain & Body | Dr....
Huberman Lab Clips



Melde dich an, um Videos mit "Mag ich" zu bewerten, zu kommentieren und um Kanäle zu abonnieren.

Anmelden

0:26 Andrew Tate || Unfazed
Filmosaur
125.737 Aufrufe • vor 2 Monaten

0:56 Life could be a dream
Asha7494
13 Mio. Aufrufe • vor 1 Jahr

2:28 ASTÉRIX AND OBÉLIX: THE MIDDLE KINGDOM Trailer...
ONE Media
2,8 Mio. Aufrufe • vor 2 Monaten

4:34 Andrew Tate vs The Matrix | Operation - Escape The Matri...
KarboTopia
362.902 Aufrufe • vor 3 Wochen

1:42:54 The System of Money | Documentary | Money Creatio...
Moconomy
1,7 Mio. Aufrufe • vor 5 Monaten

11:50 IShowSpeed Goes Sneaker Shopping With Complex
Complex
1,8 Mio. Aufrufe • vor 2 Tagen

5:02 Creed - Adonis Creed vs Ricky Conlan
MultiFandom Scene Studios
1,7 Mio. Aufrufe • vor 1 Jahr

1:25:36 Therapist Analyses Andrew Tate's Mind (FULL)
Behind Closed Doors
268.493 Aufrufe • vor 2 Monaten

Shorts

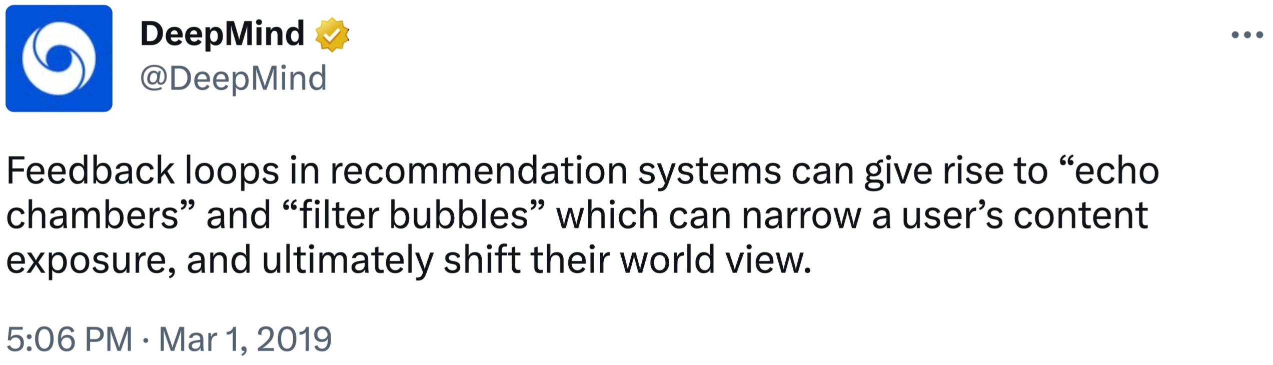
Recommendations

"Our recommendation system is built on the simple principle of helping people find the videos they **want to watch** and that will **give them value**"

Exploration vs. exploitation

https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/

Filter bubbles

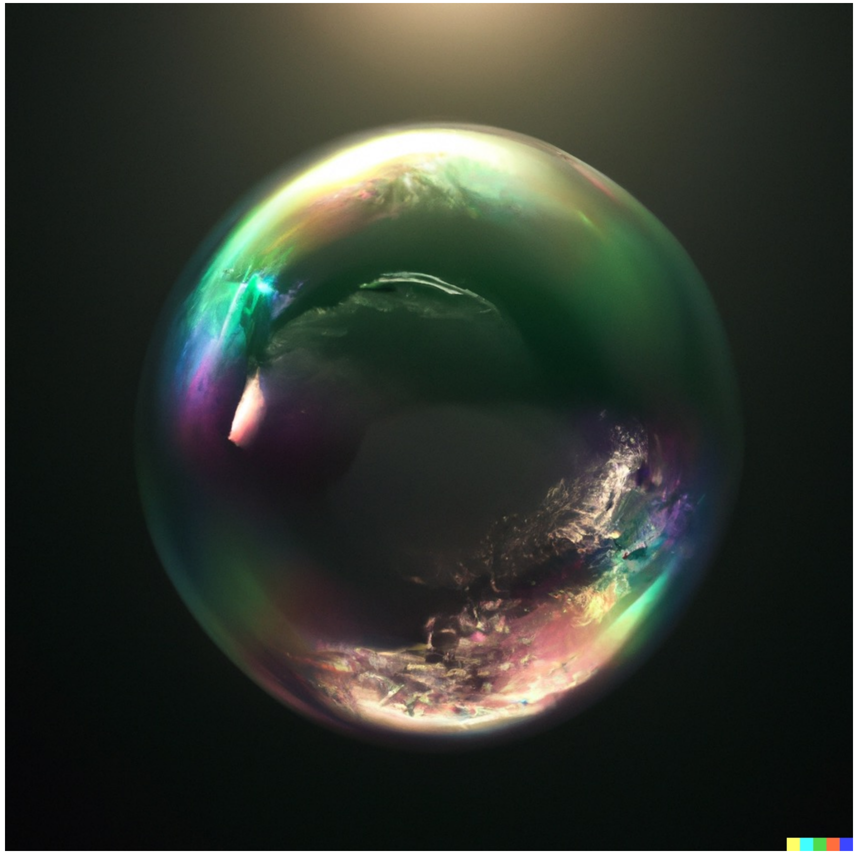


DeepMind
@DeepMind

Feedback loops in recommendation systems can give rise to “echo chambers” and “filter bubbles” which can narrow a user’s content exposure, and ultimately shift their world view.

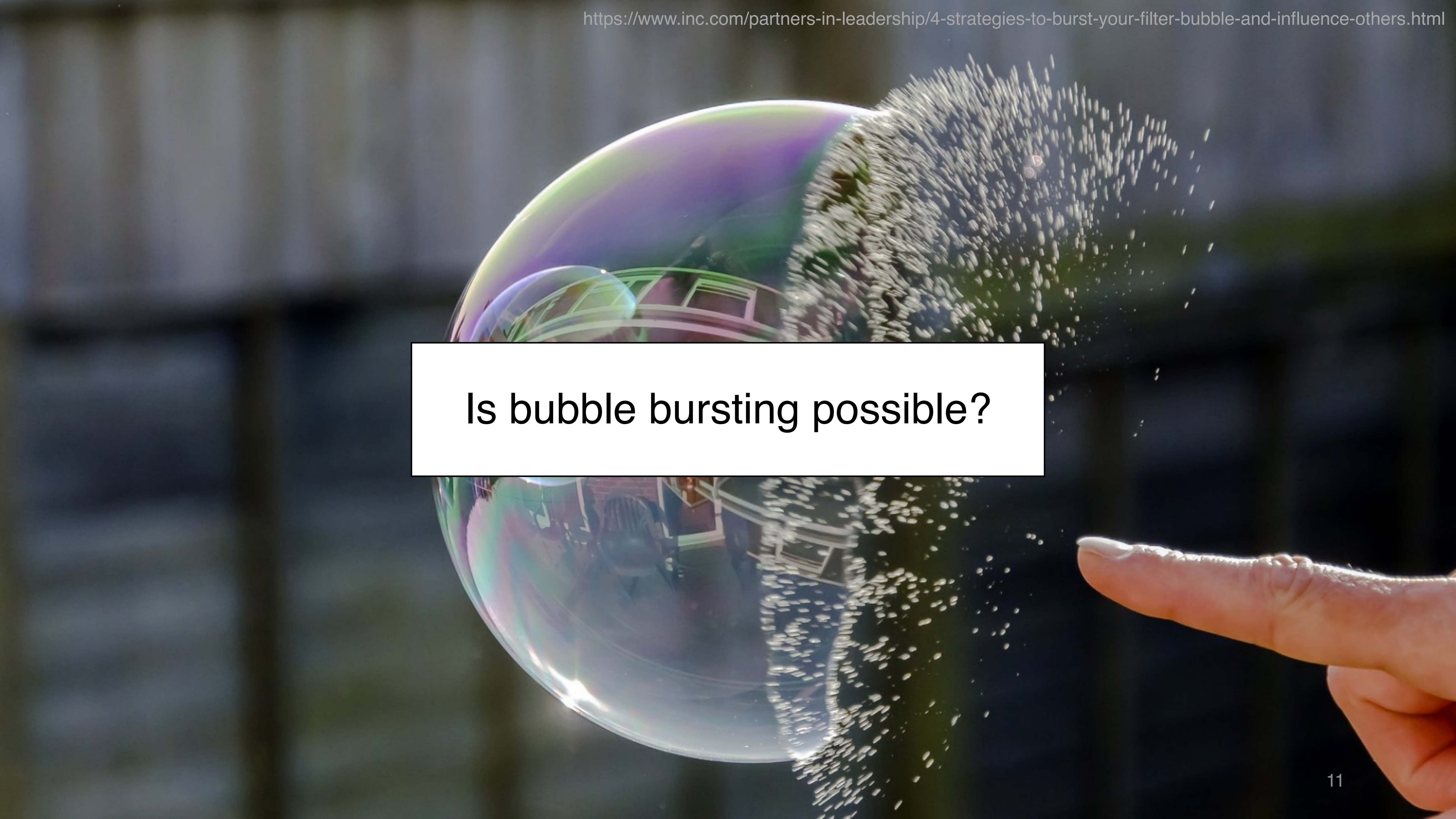
5:06 PM · Mar 1, 2019

https://twitter.com/DeepMind/status/1101514121563041792?s=20



generated by DALL-E

Is bubble bursting possible?



https://www.inc.com/partners-in-leadership/4-strategies-to-burst-your-filter-bubble-and-influence-others.html

Motivation

- Need for independent oversight of personalization behavior

generated by DALL-E

Reference study

- Hussein et al. (2020): **Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube** [1], experiment conducted in mid 2019
- This study’s experiment was conducted in March 2021

Reference study

- Filter bubbles are easily and quickly created
- “YouTube still has a long way to go to mitigate misinformation on its platform” [1]

Audits

Crowdsourcing

- using real user data
- uncontrolled environment
- hard to make comparisons

Sockpuppeting

Audits

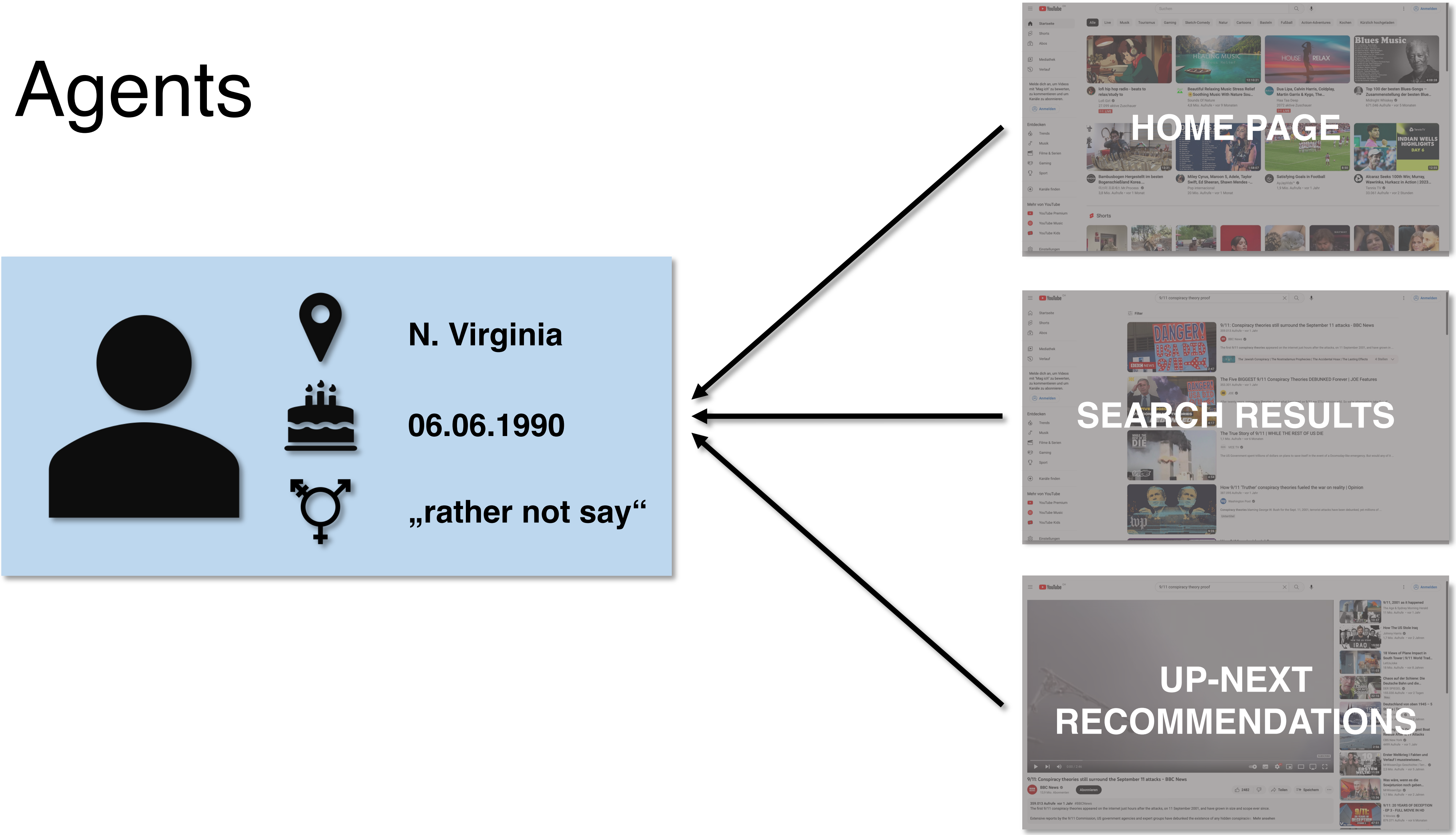
Crowdsourcing

- using real user data
- uncontrolled environment
- hard to make comparisons

Sockpuppeting

- using non-human bots
- selection of appropriate seed data

Agents



N. Virginia

06.06.1990

„rather not say“

HOME PAGE

SEARCH RESULTS

UP-NEXT RECOMMENDATIONS

Agents

- Watches videos for ≤ 30 mins
- Does **not**
  - Like
  - Subscribe
  - Comment
  - **Act human!**

Human factors

- Selective exposure
- Confirmation bias
- Dunning-Kruger effect

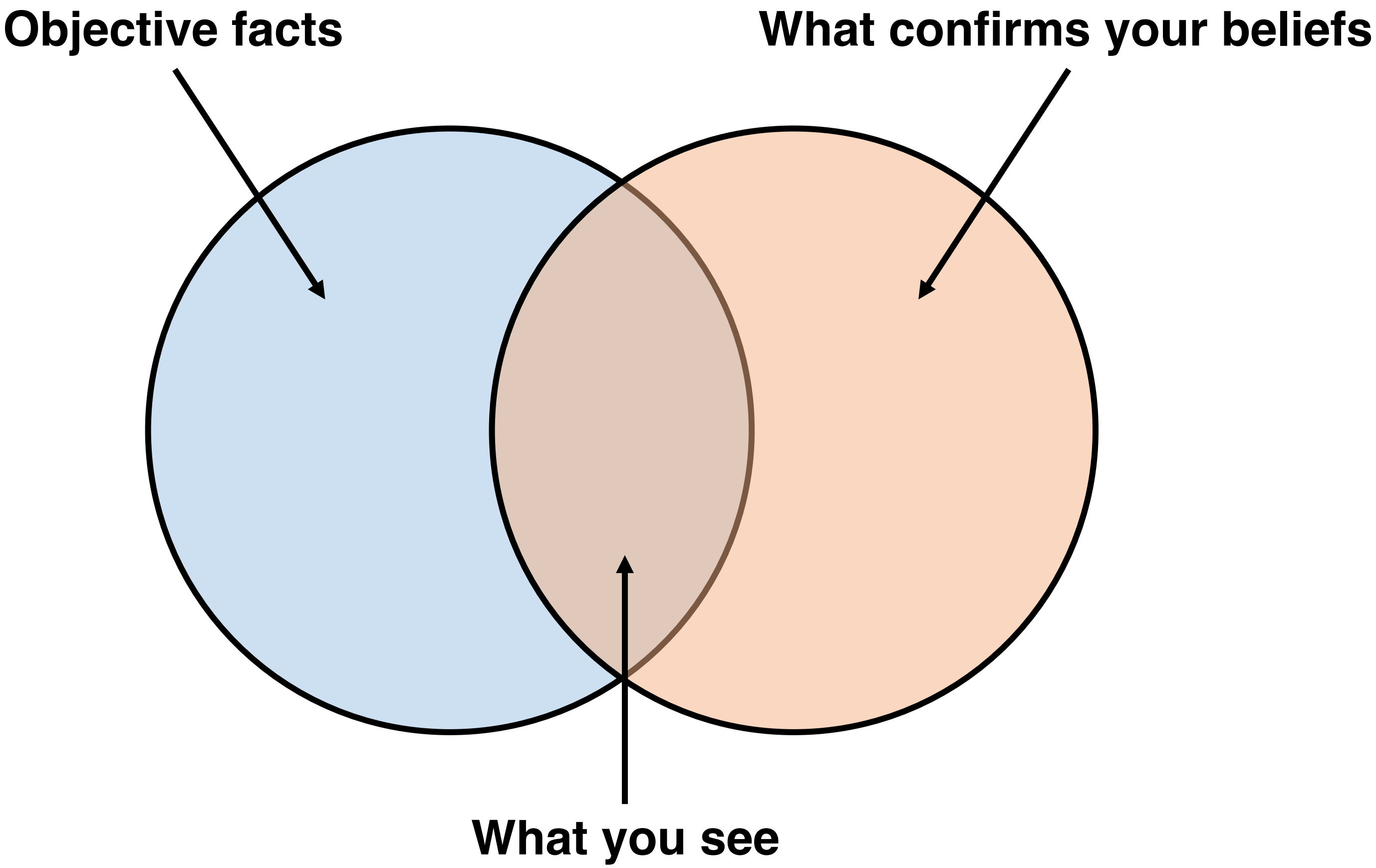
Human factors

- **Selective exposure**
- Confirmation bias
- Dunning-Kruger effect

generated by DALL-E

Human factors

- Selective exposure
- **Confirmation bias**
- Dunning-Kruger effect

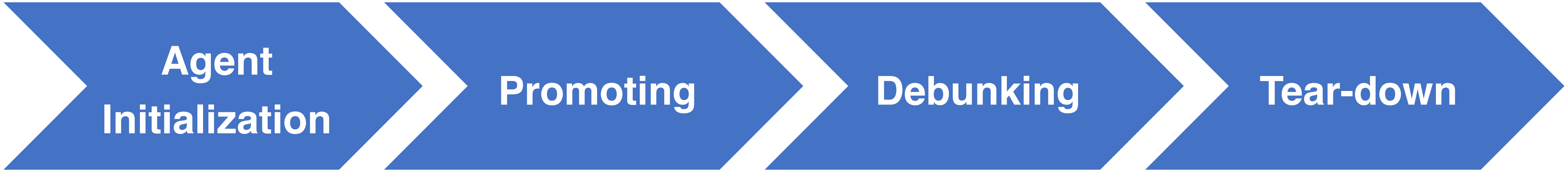


Human factors

- Selective exposure
- Confirmation bias
- **Dunning-Kruger effect**

“The less you know, the more confident you are”

Experiment



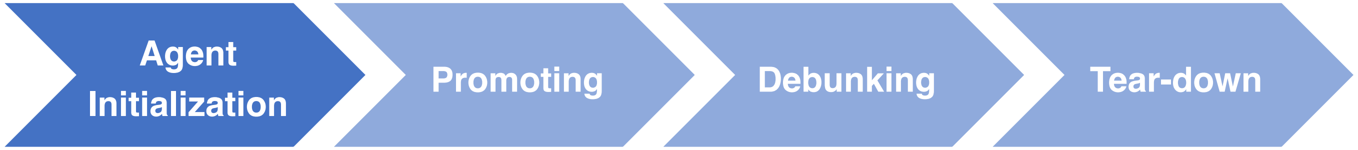
Agent Initialization

Promoting

Debunking

Tear-down

Experiment



Agent Initialization

Promoting

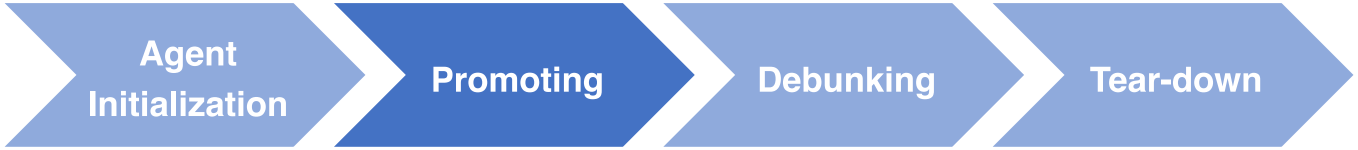
Debunking

Tear-down

Phase 0

- Most popular promoting/debunking videos (seed data)
- Search queries (e.g. “9/11 conspiracy”)
- Wait time between each query

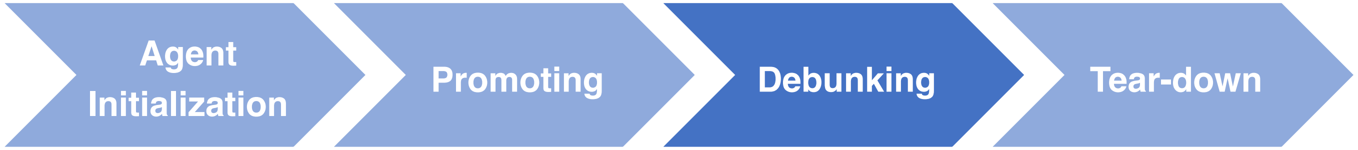
Experiment



Phase 1

- Create the filter bubble

Experiment



Agent Initialization

Promoting

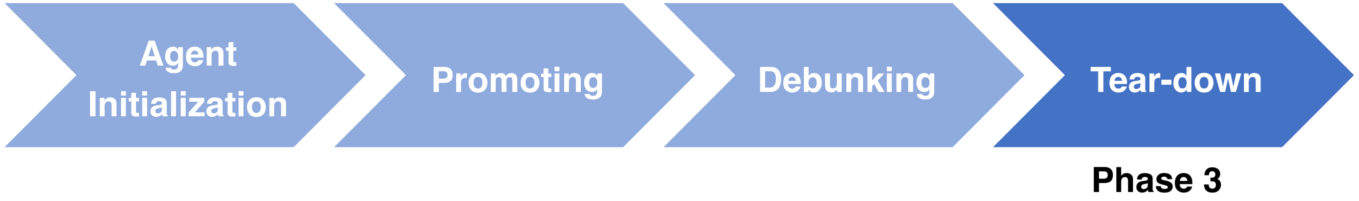
Debunking

Tear-down

Phase 2

- Burst the filter bubble

Experiment



Agent Initialization

Promoting

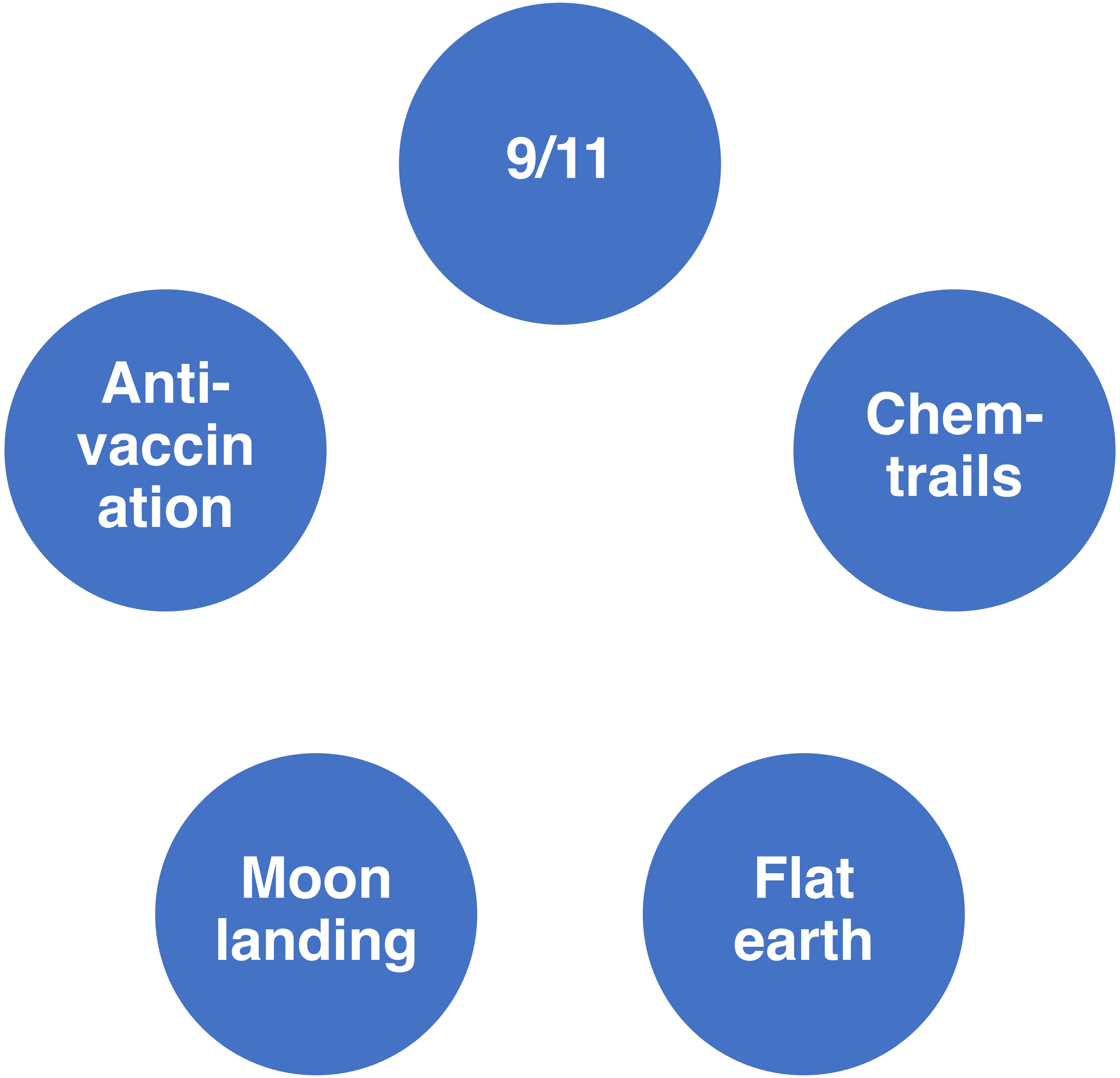
Debunking

Tear-down

Phase 3

- Clear YouTube history

Topics



9/11

Chemtrails

Flat earth

Moon landing

Anti-vaccination

Metrics

- Score x_i of a single video

$$\begin{aligned} x_i = 1 &\Leftrightarrow \quad \text{promoting} \\ x_i = 0 &\Leftrightarrow \quad \text{neutral} \\ x_i = -1 &\Leftrightarrow \quad \text{debunking} \end{aligned}$$

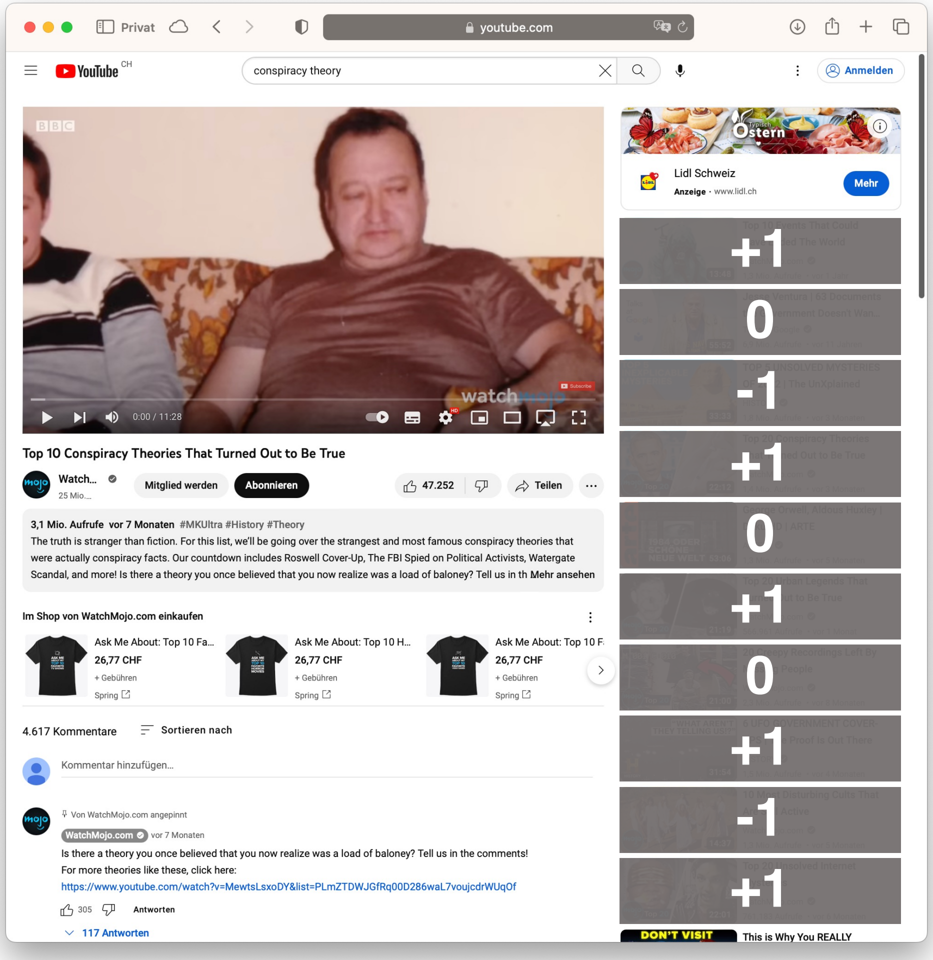
Metrics

- **N**ormalized **S**core

$$\mathrm{NS} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For recommendations

NS



Top 10 Conspiracy Theories That Turned Out to Be True

Watch... 25 Mio.... Mitglied werden Abonnieren 47.252 Teilen

3,1 Mio. Aufrufe vor 7 Monaten #MKUltra #History #Theory
The truth is stranger than fiction. For this list, we'll be going over the strangest and most famous conspiracy theories that were actually conspiracy facts. Our countdown includes Roswell Cover-Up, The FBI Spied on Political Activists, Watergate Scandal, and more! Is there a theory you once believed that you now realize was a load of baloney? Tell us in th Mehr ansehen

Im Shop von WatchMojo.com einkaufen

Ask Me About: Top 10 Fa... 26,77 CHF + Gebühren Spring

Ask Me About: Top 10 H... 26,77 CHF + Gebühren Spring

Ask Me About: Top 10 F 26,77 CHF + Gebühren Spring

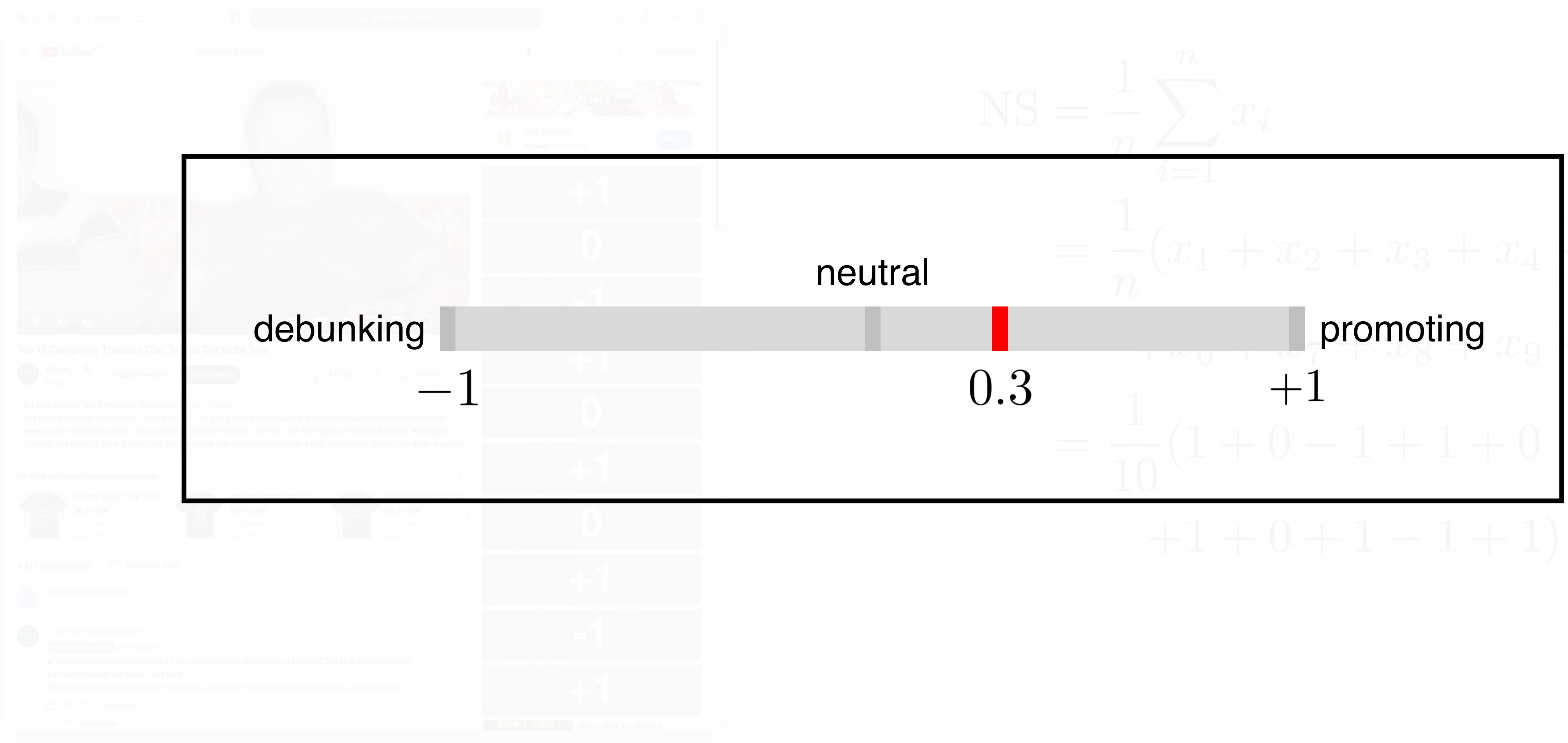
4.617 Kommentare Sortieren nach

Kommentar hinzufügen...

Von WatchMojo.com angepinnt
WatchMojo.com vor 7 Monaten
Is there a theory you once believed that you now realize was a load of baloney? Tell us in the comments!
For more theories like these, click here:
https://www.youtube.com/watch?v=MewtsLsxoDY&list=PLmZTDWJGfRq00D286waL7voujcdrWUqOf
305 Antworten
117 Antworten

$$
\begin{aligned}
\mathrm{NS} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{n}(x_1 + x_2 + x_3 + x_4 + x_5 \\
&\qquad + x_6 + x_7 + x_8 + x_9 + x_{10}) \\
&= \frac{1}{10}(1 + 0 - 1 + 1 + 0 \\
&\qquad + 1 + 0 + 1 - 1 + 1) = 0.3
\end{aligned}
$$

NS



$$\mathrm{NS} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{1}{n}(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10})$$

$$= \frac{1}{10}(1 + 0 - 1 + 1 + 0 + 1 + 0 + 1 - 1 + 1) = 0.3$$

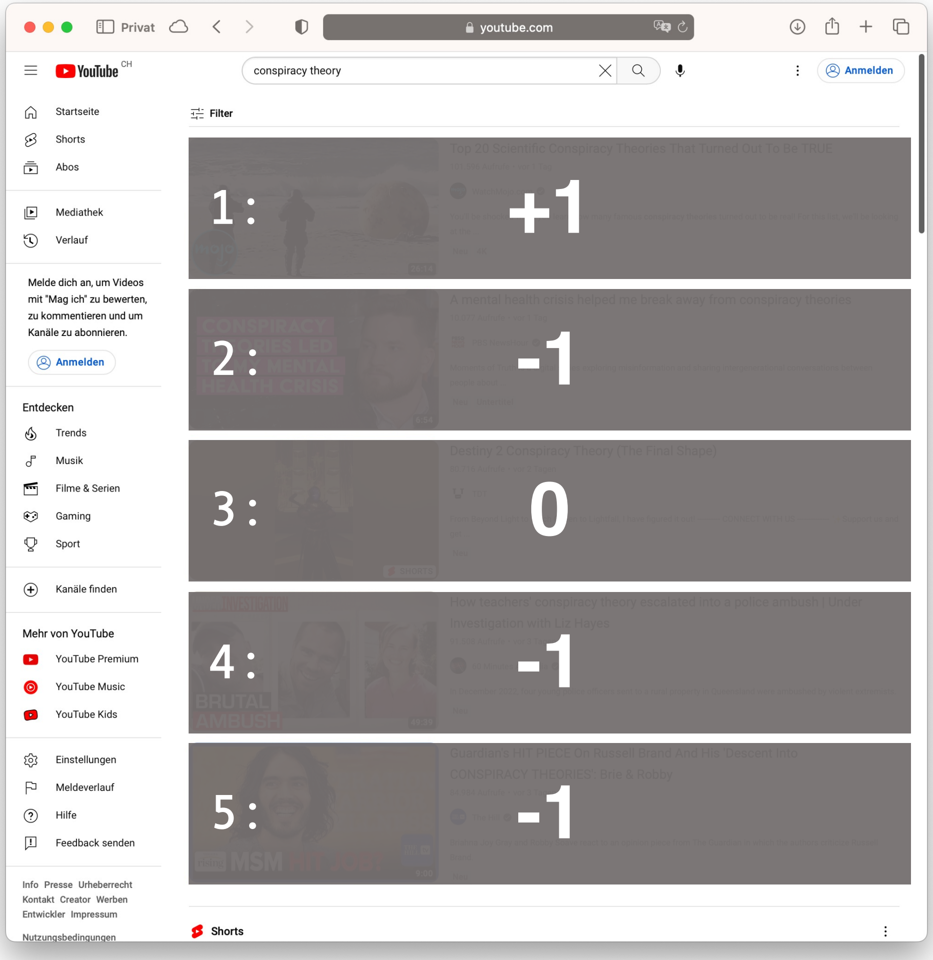
Metrics

- **Se**arch **R**esult **P**age **M**isinformation **S**core

$$\text{SERP-MS} = \frac{1}{\frac{n\cdot(n+1)}{2}} \sum_{i=1}^{n} x_i \cdot (n - \underset{\substack{\uparrow \\ \text{rank of video } i}}{r_i} + 1)$$

For search results

SERP-MS



Melde dich an, um Videos mit "Mag ich" zu bewerten, zu kommentieren und um Kanäle zu abonnieren.

1 : +1 — Top 20 Scientific Conspiracy Theories That Turned Out To Be TRUE

2 : -1 — A mental health crisis helped me break away from conspiracy theories

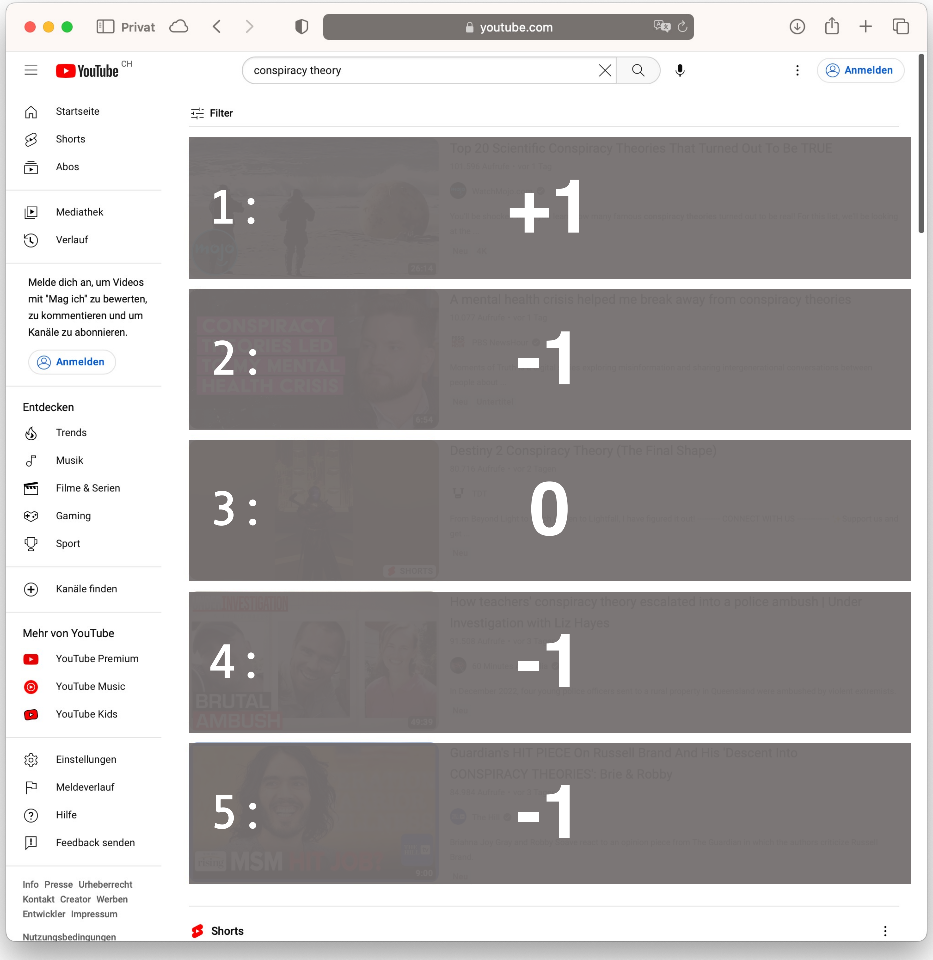
3 : 0 — Destiny 2 Conspiracy Theory (The Final Shape)

4 : -1 — How teachers' conspiracy theory escalated into a police ambush | Under Investigation with Liz Hayes

5 : -1 — Guardian's HIT PIECE On Russell Brand And His 'Descent Into CONSPIRACY THEORIES': Brie & Robby

$$
\begin{aligned}
\text{SERP-MS} &= \frac{1}{\frac{n\cdot(n+1)}{2}} \sum_{i=1}^{n} x_i \cdot (n - r_i + 1) \\
&= \frac{1}{\frac{n\cdot(n+1)}{2}} \Big(\quad x_1 \cdot (n - r_1 + 1) \\
&\qquad + x_2 \cdot (n - r_2 + 1) \\
&\qquad + x_3 \cdot (n - r_3 + 1) \\
&\qquad + x_4 \cdot (n - r_4 + 1) \\
&\qquad + x_5 \cdot (n - r_5 + 1) \Big)
\end{aligned}
$$

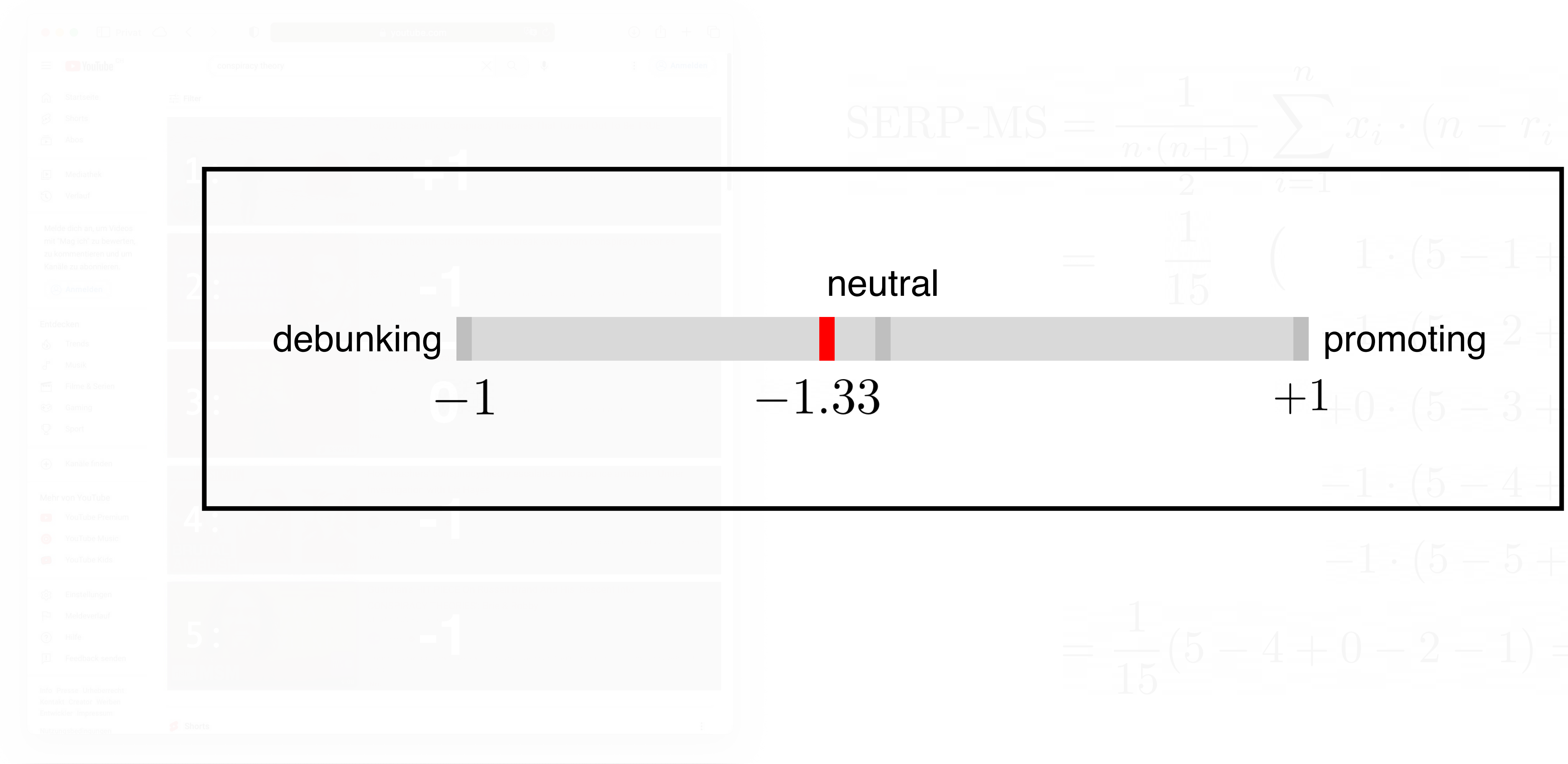
SERP-MS



Privat
youtube.com
YouTube CH
conspiracy theory
Anmelden
Startseite
Shorts
Abos
Mediathek
Verlauf
Melde dich an, um Videos mit "Mag ich" zu bewerten, zu kommentieren und um Kanäle zu abonnieren.
Anmelden
Entdecken
Trends
Musik
Filme & Serien
Gaming
Sport
Kanäle finden
Mehr von YouTube
YouTube Premium
YouTube Music
YouTube Kids
Einstellungen
Meldeverlauf
Hilfe
Feedback senden
Info Presse Urheberrecht Kontakt Creator Werben Entwickler Impressum
Nutzungsbedingungen
Filter
Top 20 Scientific Conspiracy Theories That Turned Out To Be TRUE
1 : +1
A mental health crisis helped me break away from conspiracy theories
2 : -1
Destiny 2 Conspiracy Theory (The Final Shape)
3 : 0
How teachers' conspiracy theory escalated into a police ambush | Under Investigation with Liz Hayes
4 : -1
Guardian's HIT PIECE On Russell Brand And His 'Descent Into CONSPIRACY THEORIES': Brie & Robby
5 : -1
Shorts

$$
\begin{aligned}
\text{SERP-MS} &= \frac{1}{\frac{n\cdot(n+1)}{2}} \sum_{i=1}^{n} x_i \cdot (n - r_i + 1) \\
&= \frac{1}{15} \Big(\quad 1 \cdot (5 - 1 + 1) \\
&\qquad\qquad -1 \cdot (5 - 2 + 1) \\
&\qquad\qquad +0 \cdot (5 - 3 + 1) \\
&\qquad\qquad -1 \cdot (5 - 4 + 1) \\
&\qquad\qquad -1 \cdot (5 - 5 + 1) \Big) \\
&= \frac{1}{15}(5 - 4 + 0 - 2 - 1) = -0.133
\end{aligned}
$$

SERP-MS



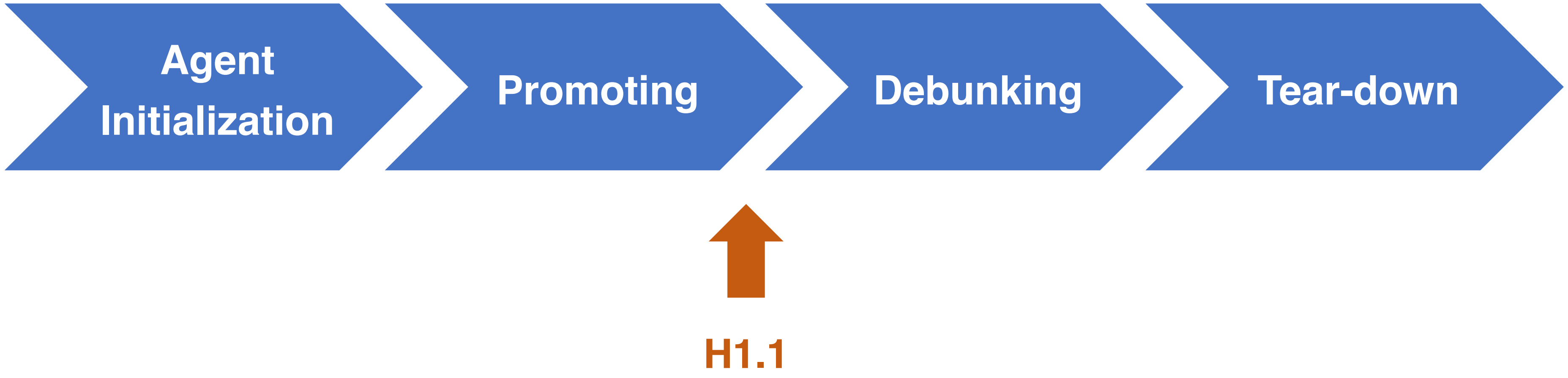
$$\text{SERP-MS} = \frac{1}{\frac{n\cdot(n+1)}{2}} \sum_{i=1}^{n} x_i \cdot (n - r_i + 1)$$

$$= \frac{1}{15} \Big(1\cdot(5-1+1) \quad -1\cdot(5-2+1) \quad +0\cdot(5-3+1) \quad -1\cdot(5-4+1) \quad -1\cdot(5-5+1) \Big)$$

$$= \frac{1}{15}(5-4+0-2-1) = -0.133$$

Hypotheses

“How has YouTube’s personalization behavior changed with regards to misinformation videos since the reference study?” [2]



Agent Initialization

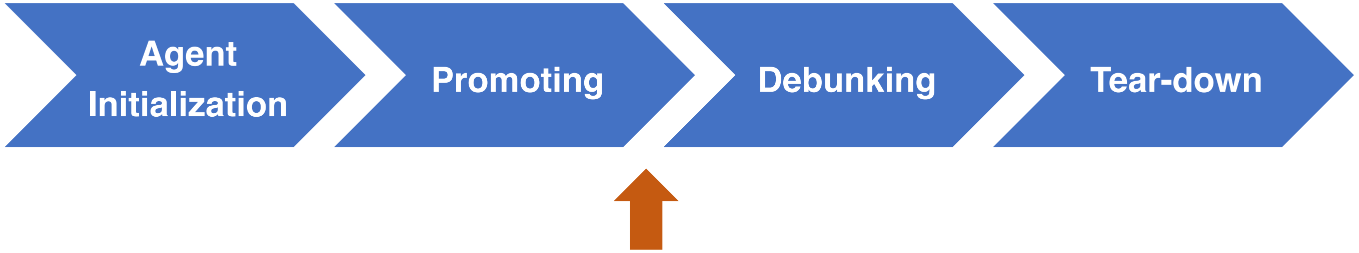
Promoting

Debunking

Tear-down

H1.1

Results



- Comparison with reference study (expecting **better**)

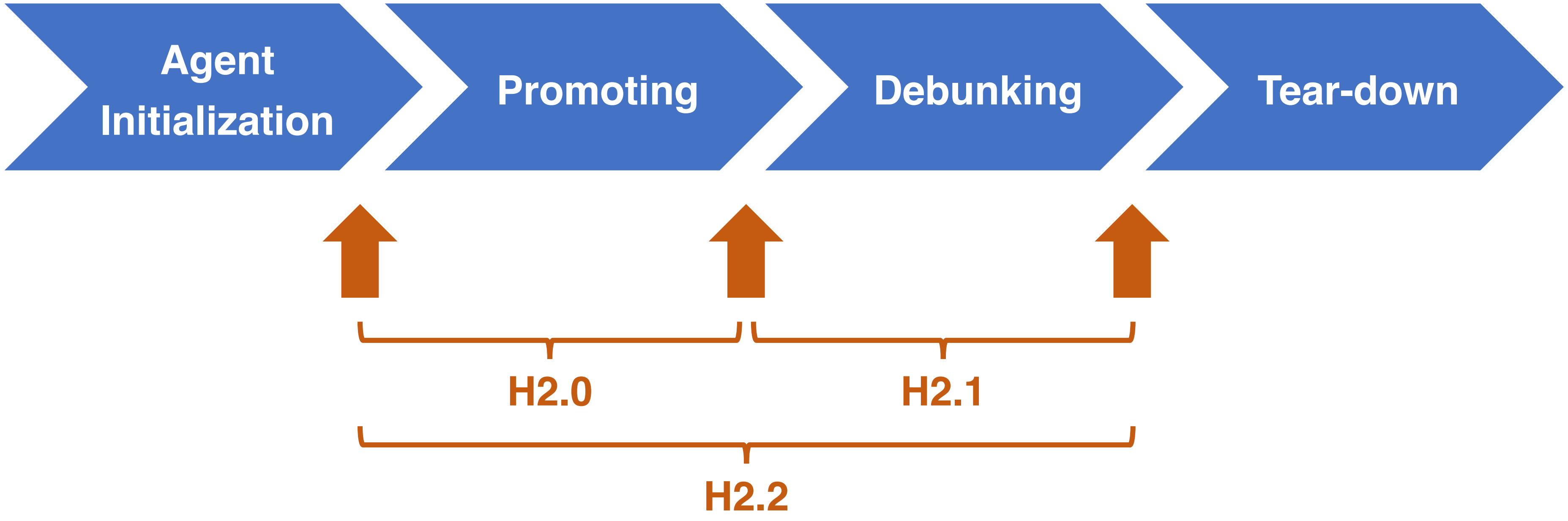
Topic	**Search results score**	**Recommendation score**
9/11	n.s.d.	n.s.d.
Chemtrails	n.s.d.	n.s.d.
Flat earth	n.s.d.	n.s.d.
Moon landing	n.s.d.	**better**
Anti-vaccination	**worse**	**worse**

Less debunking videos

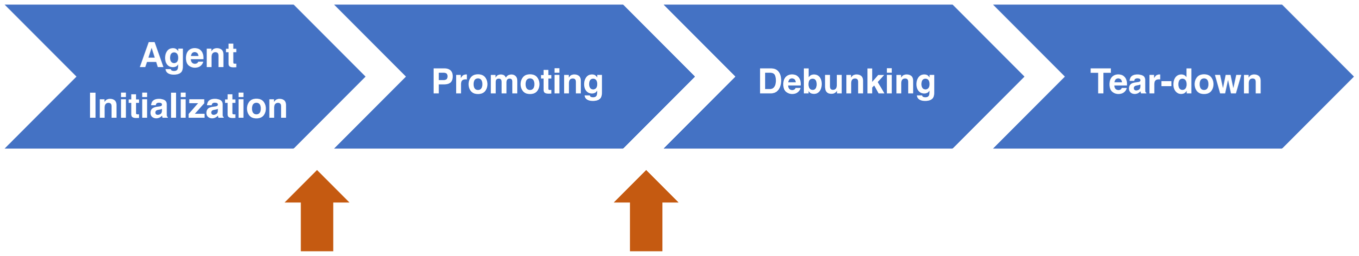
n.s.d. = not statistically significantly different

Hypotheses

“How does the effect of misinformation filter bubbles change, when debunking videos are watched?” [2]



Results



Agent Initialization

Promoting

Debunking

Tear-down

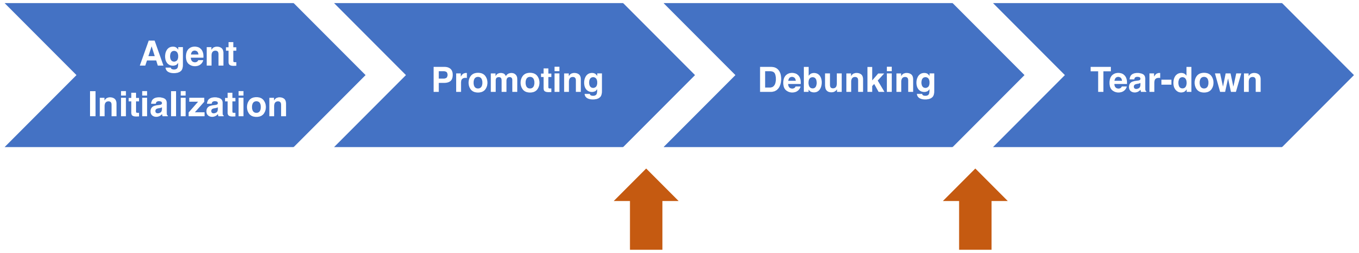
- Bubble creating behavior (expecting **worse**)

Topic	**Search results score**	**Recommendation score**
9/11	n.s.d.	**worse**
Chemtrails	n.s.d.	n.s.d.
Flat earth	**better**	n.s.d.
Moon landing	n.s.d.	n.s.d.
Anti-vaccination	n.s.d.	**worse**

Promoting videos disappear in some queries

n.s.d. = not statistically significantly different

Results

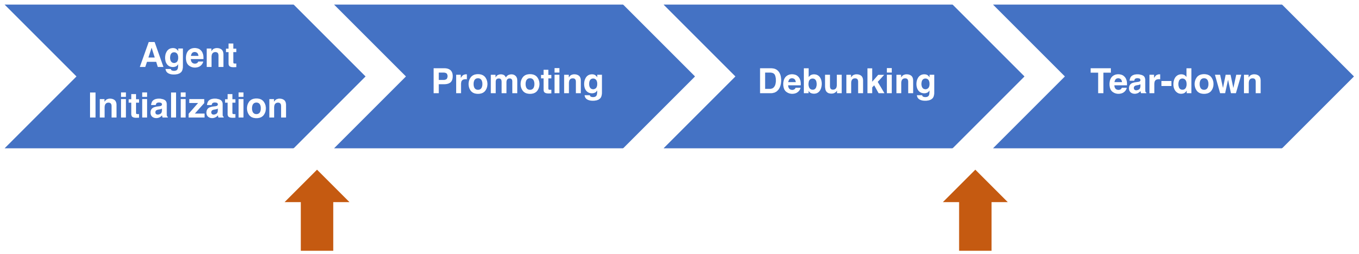


- Bubble bursting behavior (expecting **better**)

Topic	Search results score	Recommendation score
9/11	n.s.d.	**better**
Chemtrails	n.s.d.	**better**
Flat earth	n.s.d.	**better**
Moon landing	n.s.d.	n.s.d.
Anti-vaccination	**better**	**better**

n.s.d. = not statistically significantly different

Results



- Comparison to baseline (expecting **better**)

Topic	**Search results score**	**Recommendation score**
9/11	n.s.d.	n.s.d.
Chemtrails	**better**	**better**
Flat earth	**better**	**better**
Moon landing	**better**	n.s.d.
Anti-vaccination	**better**	**better**

n.s.d. = not statistically significantly different

Outlook

- Srba et al. (2023): **Auditing YouTube's Recommendation Algorithm for Misinformation Filter Bubbles** [3], continuation of this paper

Conclusion

- YouTube seems to have not fulfilled its pledges
- Bubble bursting is possible, but there are differences between topics

Discussion

- What *is* misinformation?
- How much should YouTube intervene in this matter?
- How strongly should recommendations adhere to human tendencies?
- Does YouTube treat misinformation topics differently?
- Study annotation score vs. YouTube’s “internal scoring”

References

[1] Reference study: https://dl.acm.org/doi/10.1145/3392854 (2020)

[2] Paper: https://dl.acm.org/doi/pdf/10.1145/3460231.3474241 (2021)

[3] Continuation of paper: https://arxiv.org/abs/2210.10085 (2023)

Sources

- Downward stair case: DALL-E (http://labs.openai.com), „a downward spiral into a dark dimension, digital art“ (accessed March 14, 2023)
- YouTube Recommendation System: https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/ (accessed March 13, 2023)
- Screenshot of tweet: https://twitter.com/DeepMind/status/1101514121563041792?s=20 (accessed March 14, 2023)
- Bubble: DALL-E (http://labs.openai.com), „a soap bubble, that is also a portal to a dark dimension, digital art“ (accessed March 13, 2023)
- Bubble bursting image: https://www.inc.com/partners-in-leadership/4-strategies-to-burst-your-filter-bubble-and-influence-others.html (accessed March 13, 2023)
- Computer surveillance: DALL-E (http://labs.openai.com), „dystopian 1984 themed image of computer activity being surveilled, digital art“ (accessed March 15, 2023)
- YouTube‘s pledges (October 15, 2020): https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/ (accessed March 13, 2023)
- Gender-neutral icon: https://thenounproject.com/icon/gender-neutral-147092/ (accessed March 13, 2023)
- Magnifying glass: DALL-E (http://labs.openai.com), „magnifying glass which puts only a part of a document in focus“ (accessed March 19, 2023)